WE CLAIM:

1. A method of regulating packet flow through a device having a processing fabric with at least one input port and at least one output port, a control entity connected to the at least one input port for regulating packet flow thereto, and a plurality of egress queues connected to the at least one output port for temporarily storing packets received therefrom, said method comprising:

obtaining bandwidth utilization information regarding packets received at the egress queues;

determining, from the bandwidth utilization information, a discard probability associated with each egress queue; and

providing the discard probability associated with each egress queue to the control entity, for use by the control entity in selectively transmitting packets to the at least one input port of the processing fabric.

2. A method as defined in claim 1, wherein obtaining bandwidth utilization information regarding packets received at the egress queues includes receiving said bandwidth utilization from at least one traffic management entity located between the egress queues and the at least one output port.

3. A method as claimed in claim 1, wherein each packet is made up of either a plurality of traffic bytes or a plurality of non-traffic bytes, and wherein obtaining bandwidth utilization information regarding packets received at the egress queues further includes determining, for each particular one of the at least one output port, an average number of traffic bytes received

per time unit for each egress queue connected to the particular output port.

4. A method as claimed in claim 3, wherein determining, from the bandwidth utilization information, a discard probability for a particular one of the egress queues includes:

determining an allocated traffic bandwidth for the particular egress queue;

comparing the average number of received traffic bytes for the particular egress queue to the allocated traffic bandwidth for the particular egress queue; and

if the average number of received traffic bytes for the particular egress queue is greater than the allocated traffic bandwidth for the particular egress queue, increasing the discard probability for the particular egress queue;

if the average number of received traffic bytes for the particular egress queue is less than the allocated traffic bandwidth for the particular egress queue, decreasing the discard probability for the particular egress queue.

5. A method as claimed in claim 3, wherein determining, from the bandwidth utilization information, a discard probability for a particular one of the egress queues includes:

determining an allocated traffic bandwidth for the particular egress queue;

comparing the average number of received traffic bytes for the particular egress queue to the allocated traffic bandwidth for the particular egress queue; and if the average number of received traffic bytes for the particular egress queue is greater than the allocated traffic bandwidth for the particular egress queue, setting the discard probability for the particular egress queue to the sum of a time average of previous values of the discard probability for the particular egress queue and a positive increment;

if the average number of received traffic bytes for the particular egress queue is less than the allocated traffic bandwidth for the particular egress queue, setting the discard probability for the particular egress queue to the sum of said time average of previous values of the discard probability for the particular egress queue and a negative increment.

6. A method as claimed in claim 3, wherein determining a discard probability for a particular egress queue includes:

(a) setting a temporary average number of received traffic bytes to the average number of received traffic bytes;

(b) setting a temporary discard probability equal to a time average of previous values of the discard probability for the particular egress queue;

(c) determining an allocated traffic bandwidth for the particular egress queue;

(d) comparing the temporary average number of received traffic bytes to the allocated traffic bandwidth for the particular egress queue;

(e) if the temporary average number of received traffic bytes is greater than the allocated traffic bandwidth for the particular egress queue, adding to the temporary discard probability a positive probability increment and adding to the temporary average number of received traffic bytes a negative bandwidth increment;

(f) if the temporary average number of received traffic bytes is less than the allocated traffic bandwidth for the particular egress queue, adding to the temporary discard probability a negative probability increment and adding to the temporary average number of received traffic bytes a positive bandwidth increment; and

(g) setting the discard probability for the particular egress queue to the temporary discard probability.

7. A method as defined in claim 6, further including performing steps (d), (e) and (f) a pre-determined number of times.

8. A method as defined in claim 6, further including performing steps (d), (e) and (f) until the temporary average number of received traffic bytes is within a desired range of the allocated traffic bandwidth for the particular egress queue.

9. A method as defined in claim 8, further including measuring a depth of the particular egress queue and performing steps (d), (e) and (f) until the depth of the particular egress queue is within a desired range.

10. A method as defined in claim 9, further including measuring a variability of the depth of the particular egress queue and performing steps (d), (e) and (f) until the variability of the depth of the particular egress queue is within a desired range.

11. A method as defined in claim 6, further including performing steps (d), (e) and (f) until the temporary discard probability for the particular egress queue converges to a desired precision.

12. A method as claimed in claim 6, wherein determining an allocated traffic bandwidth for the particular egress queue includes:

determining an average number of traffic bytes that would be received at the particular egress queue if the discard probability for the particular egress queue were zero; and

if the average number of traffic bytes that would be received at the particular egress queue if the discard probability for the particular egress queue were zero is greater than the allocated traffic bandwidth for the particular queue, adding a positive increment to the allocated traffic bandwidth for the particular egress queue;

if the average number of traffic bytes that would be received at the particular egress queue if the discard probability for the particular egress queue were zero is less than the allocated traffic bandwidth for the particular queue, adding a negative increment to the allocated traffic bandwidth for the particular egress queue.

13. A method as claimed in claim 6, further comprising:

determining an available traffic bandwidth for all egress queues connected to the particular output port; and

determining a total traffic bandwidth allocated for all egress queues connected to the particular output port;

wherein the step of adding a positive increment to the allocated traffic bandwidth for the particular egress queue is executed only if the total traffic bandwidth allocated for all egress queues connected to the particular output port is less than the available traffic bandwidth for all egress queues connected to the particular output port.

14. A method as claimed in claim 13, wherein determining an available traffic bandwidth for all egress queues connected to the particular output port includes:

determining a bandwidth gradient that is indicative of a rate at which the available traffic bandwidth for all egress queues connected to the particular output port is to be increased or decreased;

increasing or decreasing the available traffic bandwidth for all egress queues connected to the particular output port as a function of the bandwidth gradient.

15. A method as claimed in claim 14, wherein obtaining bandwidth utilization information regarding packets

received at the egress queues further includes determining, for each particular one of the at least one output port, an average number of non-traffic bytes received per time unit from the particular output port, and wherein determining an available traffic bandwidth for all egress queues connected to the particular output port further includes:

determining a total link capacity available for all the egress queues connected to the particular output port;

setting a maximum available traffic bandwidth to the difference between said total link capacity and said average number of non-traffic bytes;

wherein the available traffic bandwidth for all egress queues connected to the particular output port is bounded above by the maximum available traffic bandwidth.

16. A method as claimed in claim 15, wherein determining the average number of traffic bytes that would be received at the particular egress queue if the discard probability for the particular egress queue were zero includes evaluating a function of the average number of traffic bytes received per time unit for the particular egress queue and the time average of previous values of the discard probability for the particular egress queue.

17. A method as claimed in claim 16, wherein the function is the quotient between (i) the average number of traffic bytes received per time unit for the particular egress queue and (ii) the difference between unity and the time average of previous values of the discard probability for the particular egress queue.

18. A method as claimed in claim 6, further comprising:

determining an average number of traffic bytes that would be received at the particular egress queue if the discard probability for the particular egress queue were zero; and

performing steps (d), (e) and (f) at least twice;

wherein the positive bandwidth increment is a first fraction of average number of traffic bytes that would be received at the particular egress queue if the discard probability for the particular egress queue were zero, said first fraction decreasing with subsequent executions of step (f); and

wherein the negative bandwidth increment is a second fraction of average number of traffic bytes that would be received at the particular egress queue if the discard probability for the particular egress queue were zero, said second fraction decreasing with subsequent executions of step (e).

19. A method as claimed in claim 18, wherein the positive probability increment has a value that decreases with subsequent executions of step (e) and wherein the negative probability increment has a value that decreases with subsequent executions of step (f).

20. A method as defined in claim 14, wherein obtaining bandwidth utilization information regarding packets received at the egress queues includes determining, for each particular one of the at least one output port, an average idle time between successive packets received from the particular output port.

21. A method as claimed in claim 20, wherein determining a bandwidth gradient includes:

comparing the average idle time between successive packets received from the particular output port to a first threshold; and

if the average idle time between successive packets received from the particular output port is below the first threshold, setting the bandwidth gradient to indicate a first rate of decrease in the available traffic bandwidth for all egress queues connected to the particular output port.

22. A method as claimed in claim 21, further comprising:

comparing the average idle time between successive packets received from the particular output port to a second threshold less than the first threshold; and

if the average idle time between successive packets received from the particular output port is below the second threshold, setting the bandwidth gradient to indicate a second rate of decrease in the available traffic bandwidth for all egress queues connected to the particular output port, wherein said second rate of decrease is greater than said first rate of decrease.

23. A method as claimed in claim 22, further comprising:

comparing the average idle time between successive packets received from the particular output port to a third threshold; and

if the average idle time between successive packets received from the particular output port is above the third threshold, setting the bandwidth gradient to indicate a rate of increase in the available traffic bandwidth for all egress queues connected to the particular output port.

24. A method as claimed in claim 23, further comprising:

determining a degree of memory utilization within the plurality of egress queues; and

programming at least one of the first, second and third threshold as a function of said degree of memory utilization.

25. A method as claimed in claim 1, wherein the at least one output port of the processing fabric is a plurality of output ports and wherein each of the plurality of output ports is connected to a respective one of the plurality of egress queues.

26. A method as claimed in claim 1, wherein at least one of the at least one output port of the processing fabric is connected to a respective plurality of the plurality of egress queues.

27. A method as claimed in claim 1, wherein providing the discard probability associated with each egress queue to the control entity is executed on a programmable basis.

28. A method as claimed in claim 1, further comprising:

recording the discard probability associated with each egress queue at selected times;

detecting whether a change of at least a pre-determined magnitude has occurred in the discard probability associated with at least one of the egress queues;

wherein providing the discard probability associated with a particular one of the egress queues to the control entity is executed only if a change of at least the pre-determined magnitude has been detected in the discard probability associated with the particular egress queue.

29. A method as claimed in claim 1, further comprising:

recording the discard probability associated with each egress queue at selected times;

detecting whether a change of at least a pre-determined magnitude has occurred in the discard probability associated with at least one of the egress queues;

wherein providing the discard probability associated with a particular one of the egress queues to the control entity is executed either (i) if a change of at least the pre-determined magnitude has been detected in the discard probability associated with the particular egress queue; or (ii) after a pre-determined amount of time regardless of whether or not a change of at least the pre-determined magnitude has been detected in the discard probability associated with the particular egress queue.

30. A method as claimed in claim 1, further comprising:

for each received packet, the control entity determining an egress queue for which the received packet is destined and either transmitting or not transmitting the received packet to the processing fabric on the basis of the discard probability associated with the egress queue for which the received packet is destined.

31. A method as claimed in claim 30, wherein either transmitting or not transmitting the received packet to the processing fabric on the basis of the discard probability associated with the egress queue for which the received packet is destined includes:

generating a random number for the received packet;

comparing the random number to the discard probability associated with the egress queue for which the received packet is destined; and

transmitting or not transmitting the received packet to the processing fabric on the basis of the comparison.

32. A method as claimed in claim 31, wherein not transmitting a received packet includes discarding the packet.

33. A method as claimed in claim 31, wherein not transmitting a received packet includes marking the packet as discardable.

34. A method as claimed in claim 31, wherein not transmitting a received packet includes storing the received packet in a memory location and marking the received packet as discardable, and wherein transmitting a received packet includes transmitting only those packets not marked as discardable.

35. A method as claimed in claim 34, wherein not transmitting a received packet further includes:

determining whether there exists a condition of reduced congestion at the egress queues; and if there exists a condition of reduced congestion at the egress queues, determining whether the memory location needs to be used to store another packet and, if not, unmarking the packet as discardable.

36. A computer-readable storage medium containing program instructions for causing execution in a computing device of a method as defined in claim 1.

37. A drop probability evaluation module for use in a device having (i) a processing fabric with at least one input port and at least one output port; (ii) a control entity connected to the at least one input port for regulating packet flow thereto; and (iii) a plurality of egress queues connected to the at least one output port for temporarily storing packets received therefrom, said drop probability evaluation module comprising:

means for obtaining bandwidth utilization information regarding packets received at the egress queues;

means for determining, from the bandwidth utilization information, a discard probability associated with each egress queue; and

means for providing the discard probability associated with each egress queue to the control entity, for use by the control entity in selectively transmitting packets to the at least one input port of the processing fabric.

38. A drop probability evaluation module for use in a device having (i) a processing fabric with at least one input port and at least one output port; (ii) a control

entity connected to the at least one input port for regulating packet flow thereto; and (iii) a plurality of egress queues connected to the at least one output port for temporarily storing packets received therefrom, said drop probability evaluation module including:

an allocation processing entity, for determining an allocated traffic bandwidth for each of the egress queues; and

a probability processing entity in communication with the allocation processing entity, said probability processing entity being adapted to receive the allocated traffic bandwidth for each of the egress queues from the allocation processing entity and also adapted to receive an average number of received traffic bytes for each of the egress queues from an external entity, the probability processing entity being operable to:

compare the average number of received traffic bytes for each particular one of the egress queues to the allocated traffic bandwidth for the particular egress queue; and

set the discard probability for the particular egress queue to the sum of a time average of previous values of the discard probability for the particular egress queue and either a positive or a negative increment, depending on whether the average number of received traffic bytes for the particular egress queue is greater or less than the allocated traffic bandwidth for the particular egress queue.

39. A computer-readable storage medium containing a program element for execution by a computing device to

implement the drop probability evaluation module of claim 38.

40. An apparatus, comprising:

a processing fabric having at least one input port and at least one output port, the processing fabric being adapted to process packets received from the at least one input port and forward processed packets to the at least one output port;

a plurality of egress queues, each connected to a corresponding one of the at least one output port of the processing fabric, each egress queue being adapted to (i) temporarily store packets received from the corresponding output port of the processing fabric and (ii) determine bandwidth utilization information on the basis of the packets received at the egress queues;

a drop probability evaluation module connected to the egress queues, said drop probability evaluation entity being adapted to determine a discard probability associated with each of the egress queues on the basis of the bandwidth utilization information; and

a packet acceptance unit connected to the at least one input port of the processing fabric and to the drop probability evaluation module, the packet acceptance entity being adapted to (i) receive packets destined for the at least one output port of the processing fabric; (ii) identify an egress queue associated with each received packet; and (iii) on the basis of the discard probability associated with the egress queue associated with each received packet, either transmit or not transmit said received packet to one of the at least one input port of the processing fabric.

41. Apparatus as claimed in claim 40, wherein the at least one output port is a plurality of output ports, the apparatus further comprising:

a plurality of output line cards, each output line card connected to a distinct subset of the plurality of output ports of the processing fabric;

wherein a portion of the drop probability evaluation module is provided on each of the output line cards;

wherein the portion of the drop probability evaluation module provided on a particular one of the output line cards is the portion of the drop probability evaluation module connected to those egress queues that are connected to the subset of the plurality of output ports of the processing fabric to which the particular output line card is connected.

42. Apparatus as claimed in claim 41, wherein the at least one input port is a plurality of input ports further comprising:

a plurality of input line cards, each input line card being connected to a distinct subset of the plurality of input ports of the processing fabric;

wherein a portion of the packet acceptance unit is provided on each of the input line cards.

43. Apparatus as defined in claim 40, wherein the processing fabric is a switch fabric.

44. A method of regulating packet flow through a device having a processing fabric with at least one input port and at least one output port, a control entity connected

to the at least one input port for regulating packet flow thereto, and a plurality of egress queues connected to the at least one output port for temporarily storing packets received therefrom, each packet having a corresponding priority selected from a group of priorities, said method comprising:

obtaining bandwidth utilization information regarding packets received at the egress queues;

determining, from the bandwidth utilization information, a discard probability associated with each of the egress queues and each of the priorities; and

providing the discard probability associated with each egress queue and priority to the control entity, for use by the control entity in selectively transmitting packets to the at least one input port of the processing fabric.

45. A method as claimed in claim 44, further comprising:

for each received packet, the control entity determining an egress queue for which the received packet is destined and the priority of the packet and either transmitting or not transmitting the received packet to the processing fabric on the basis of the discard probability associated with the egress queue for which the received packet is destined and the priority of the packet.

46. A method of regulating packet flow through a device having an ingress entity, an egress entity, a processing fabric between the ingress entity and the egress entity, and a control entity adapted to process packets prior to

transmission thereof to the ingress entity, said method comprising:

obtaining congestion information regarding packets received at the egress entity; and

providing the congestion information to the control entity, for use by the control entity in processing packets prior to transmission thereof to the ingress entity.

47. A method as defined in claim 46, further comprising:

for each packet received at the control entity, either transmitting or not transmitting the received packet to the ingress entity, on the basis of the congestion information.

48. A method as defined in claim 47, wherein not transmitting the received packet to the ingress entity includes discarding the received packet.

49. A method as defined in claim 47, wherein not transmitting the received packet to the ingress entity includes storing the packet in a memory location.

50. A method as defined in claim 47, wherein not transmitting the received packet to the ingress entity includes rerouting the packet along an alternate route.

51. A method as defined in claim 46, further comprising:

for each packet received at the control entity, either marking or not marking the received packet prior to transmission to the ingress entity, on the basis of the congestion information.

52. A method as defined in claim 51, further including detecting congestion by receiving marked packets at the egress entity.

53. A method as defined in claim 46, wherein obtaining congestion information regarding packets received at the egress entity includes determining a discard probability.

54. A method as defined in claim 53, further including:

generating a quantity for each packet received at the control entity;

comparing the quantity to the discard probability; and

either transmitting or not transmitting the received packet to the ingress entity on the basis of the outcome of the comparing step.

55. A method as defined in claim 54, wherein the quantity is a random number.

56. A method as defined in claim 46, wherein the congestion information includes bandwidth utilization information.

57. A method as defined in claim 46, wherein the egress entity includes a plurality of egress queues and wherein the congestion information includes an occupancy of each of the egress queues.

58. A method as defined in claim 57, wherein the egress entity includes a plurality of egress queues and wherein

the congestion information includes a variability in the occupancy of each of the egress queues.